Modification of Protein Glycosylation in Methylotrophic Yeast

Field of the Invention

The present invention relates to methods and genetically engineered methylotrophic yeast strains for producing glycoproteins with mammalian-like glycosylation. The present invention also relates to vectors useful for generating methylotrophic yeast strains capable of producing glycoproteins with mammalian-like glycosylation. Glycoproteins produced from the genetically engineered methylotrophic yeast strains are also provided.

Background of the Invention

The methylotrophic yeasts including *Pichia pastoris* have been widely used for production of recombinant proteins of commercial or medical importance. However, production and medical applications of some therapeutic glycoproteins can be hampered by the differences in the protein-linked carbohydrate biosynthesis between these yeasts and the target organism such as a mammalian or human subject.

Protein N-glycosylation originates in the endoplasmic reticulum (ER), where an N-linked oligosaccharide ($Glc_3Man_9GlcNAc_2$) assembled on dolichol (a lipid carrier intermediate) is transferred to the appropriate Asn of a nascent protein. This is an event common to all eukaryotic N-linked glycoproteins. The three glucose residues and one specific α-1,2-linked mannose residue are removed by specific glucosidases and an α-1,2-mannosidase in the ER, resulting in the core oligosaccharide structure, $Man_8GlcNAc_2$. The protein with this core sugar structure is transported to the Golgi apparatus where the sugar moiety undergoes various modifications. There are significant differences in the modifications of the sugar chain in the Golgi apparatus between yeast and higher eukaryotes.

In mammalian cells, the modification of the sugar chain proceeds via 3 different pathways depending on the protein moiety to which it is added. That is, (1) the

core sugar chain does not change; (2) the core sugar chain is changed by the addition of the N-acetylglucosamine-1-phosphate moiety (GlcNAc-1-P) from UDP-N-acetyl glucosamine (UDP-GlcNAc) to the 6-position of mannose in the core sugar chain, followed by removal of the GlcNAc moiety to form an acidic sugar chain in the glycoprotein; or (3) the core sugar chain is first converted into $Man_5GlcNAc_2$ as a result of the removal of 3 mannose residues by mannosidase I; and $Man_5GlcNAc_2$ is further modified by the addition of GlcNAc and the removal of two more mannose residues, followed by the sequential addition of GlcNAc, galactose (Gal), and N-acetylneuraminic acid (also called sialic acid (NeuNAc)) to form various hybrid or complex sugar chains (R. Kornfeld and S. Kornfeld, *Ann. Rev. Biochem.* 54: 631-664, 1985; Chiba et al *J. Biol. Chem.* 273: 26298-26304, 1998).

In yeast, the $Man_8GlcNAc_2$ glycans are not trimmed. The modification of the sugar chain in the Golgi apparatus involves a series of additions of mannose residues by different mannosyltransferases ("outer chain" glycosylation). The structure of the outer chain glycosylation is specific to the organisms, typically with more than 50 mannose residues in *S. cerevisiae*, and most commonly with structures smaller than $Man_{14}GlcNAc_2$ in *Pichia pastoris*. This yeast-specific outer chain glycosylation of the high mannose type is also denoted as hyperglycosylation or hypermannosylation.

Glycosylation is crucial for correct folding, stability and bioactivity of proteins. In the human body, glycosylation is partially responsible for the pharmacokinetic properties of a protein, such as tissue distribution and clearance from the blood stream. In addition, glycan structures can be involved in antigenic responses. For example, the presence of α-galactose on glycoproteins is the main reason for the immune reaction against xenografts from pig (Chen et al., *Curr Opin Chem Biol,* 3(6):650-658, 1999), while the immune reaction against glycoproteins from yeast is mainly due to the presence of α-1,3-mannose, β-linked mannose and/or phosphate residues in either a phosphomono- or phosphodiester linkage (Ballou, C.E., *Methods Enzymol,* 185:440-470, 1990; Yip et al., *Proc Natl Acad Sci USA,* 91(7):2723-2727, 1994).

Hyperglycosylation is often undesirable since it leads to heterogeneity of a recombinant protein product in both carbohydrate composition and molecular weight, which may complicate purification of the protein. The specific activity (units/weight) of hyperglycosylated enzymes can be lowered by the increased portion of carbohydrate. In addition, the outer chain glycosylation is often strongly immunogenic which may be undesirable in a therapeutic application. Moreover, the large outer chain sugar can mask the immunogenic determinants of a therapeutic protein. For example, the influenza neuraminidase (NA) expressed in *P. pastoris* is glycosylated with N-glycans containing up to 30-40 mannose residues. The hyperglycosylated NA has a reduced immunogenicity in mice, as the variable and immunodominant surface loops on top of the NA molecule are masked by the N-glycans (Martinet et al. *Eur J. Biochem.* 247: 332-338, 1997).

Therefore, it is desirable to genetically engineer methylotrophic yeast strains which produce recombinant glycoproteins having carbohydrate structures that resemble mammalian (e.g., human) carbohydrate structures.

Summary of the Invention

The present invention is directed to genetically engineered methylotrophic yeast strains and methods for producing glycoproteins with mammalian-like N-glycans. The present invention is also directed to vectors and kits useful for generating the genetically engineered methylotrophic yeast strains capable of producing glycoproteins with mammalian-like N-glycans.

The term "methylotrophic yeast" as used herein includes, but is not limited to, yeast strains capable of growing on methanol, such as yeasts of the genera *Candida, Hansenula, Torulopsis,* and *Pichia.*

In one embodiment, the present invention provides a genetically engineered methylotrophic yeast strain which produces glycoproteins having a mammalian-like N-glycan structure, characterized by having five or fewer mannose residues and at least one N-acetylglucosamine residue (GlcNAc) which is linked to the core mannose-containing structure and to a terminal galactose residue.

In a preferred embodiment, the present invention provides a genetically engineered methylotrophic yeast strain which produces glycoproteins having the mammalian-like N-glycan structure, $GalGlcNAcMan_5GlcNAc_2$.

According to the present invention, the methylotrophic yeast strain which produces glycoproteins having $GalGlcNAcMan_5GlcNAc_2$ is genetically engineered to express an α-1,2-mannosidase or a functional part thereof, an N-acetylglucosaminyltransferase I (or GnTI) or a functional part thereof, and a β-1,4-galactosyltransferase (GalT) or a functional part thereof. Preferably, the methylotrophic yeast strain is also genetically engineered such that the genomic OCH1 gene is inactivated.

The α-1,2-mannosidase or a functional part thereof for expression in a genetically engineered methylotrophic yeast strain can be of an origin of any species, including mammalian species such as murine, rabbit or human, and fungal species such as *Aspergillus*, or *Trichoderma reesei*. A preferred α-1,2-mannosidase for use in the present invention is the *Trichoderma reesei* α-1,2-mannosidase. Preferably, the α-1,2-mannosidase or a functional part thereof is targeted to a site in the secretory pathway where its substrate, $Man_8GlcNAc_2$, is available. More preferably, the α-1,2-mannosidase or a functional part thereof is genetically engineered to contain an ER-retention signal and is targeted to the ER. A preferred ER-retention signal is the peptide, HDEL (SEQ ID NO: 1).

The GnTI or a functional part thereof for expression in a genetically engineered methylotrophic yeast strain can be of an origin of any species, including rabbit, rat, human, plants, insects, nematodes and protozoa such as *Leishmania tarentolae*. A preferred GnTI for use in the present invention is the human GnTI as set forth in SEQ ID NO: 13. Preferably, the GnTI or a functional part thereof is targeted to a site in the secretory pathway where its substrate, $Man_5GlcNAc_2$, is available. More preferably, the GnTI or a functional part thereof is genetically engineered to contain a Golgi-retention signal and is targeted to the Golgi apparatus. A preferred a Golgi-

retention signal is the peptide as set forth in SEQ ID NO: 11, composed of the first 100 amino acids of the *Saccharomyces cerevisiae* Kre2 protein.

The GalT or a functional part thereof for expression in a genetically engineered methylotrophic yeast strain can be of an origin of any species, including human, plants (e.g. *Arabidopsis thaliana*), insects (e.g. *Drosophila melanogaster*). A preferred GalT for use in the present invention is the human GalTI as set forth in SEQ ID NO: 21. Preferably, the GalT or a functional part thereof is genetically engineered to contain a Golgi-retention signal and is targeted to the Golgi apparatus. A preferred Golgi-retention signal is the peptide as set forth in SEQ ID NO: 11, composed of the first 100 amino acids of the *Saccharomyces cerevisiae* Kre2 protein.

A methylotrophic yeast strain can be genetically engineered to express the above desired enzymes by introducing into the strain nucleotide sequences coding for these enzymes by way of, e.g., transformation. Preferably, the coding sequences are provided in vectors, each sequence placed in an operable linkage to a promoter sequence and a 3' termination sequence that are functional in the yeast strain. The vectors or linear fragments thereof are then transformed into the strain.

According to a preferred embodiment of the present invention, the methylotrophic yeast strain is also genetically engineered such that the genomic OCH1 gene is disrupted. Gene disruption can be achieved by homologous recombination between the genomic OCH1 sequence and the OCH1 sequence(s) in a knock-out vector.

In a further aspect, the present invention provides vectors useful for generating methylotrophic yeast strains which produces glycoproteins having a mammalian-like N-glycan structure.

In one embodiment, the present invention provides a "knock-in" vector which contains a nucleotide sequence coding for an enzyme to be expressed, i.e., an α-1,2-mannosidase, a GnTI, a GalT, or a functional part of any of these proteins. The coding sequence can be placed in an operable linkage to a promoter and a 3' termination sequence that are functional in the host methylotrophic yeast for expression of the encoded protein. Two or more coding sequences can be placed in the same vector for simultaneous transformation into a methylotrophic yeast strain. Preferably, the vector

also includes a selectable marker gene for convenient selection of transformants. A knock-in vector can be an integrative vector or a replicative vector.

In another embodiment, the present invention provides an inactivation vector (or a "knock-out" vector) which, when introduced into a methylotrophic yeast strain, inactivates or disrupts the genomic OCH1 gene.

The OCH1 knock-out vector can include a selectable marker gene, which is operably linked, at both its 5' and 3' end, to OCH1 sequences of lengths sufficient to mediate double homologous recombination with the genomic OCH1 gene. Alternatively, an OCH1 inactivation vector can include a portion of the OCH1 gene to be disrupted, which portion encodes none or an inactive fragment of the OCH1 protein, and a selectable marker gene. The OCH1 portion is not in an operable linkage to any known promoter sequence and can, upon transformation of linear fragments of the vector, integrate into the genomic OCH1 locus by single homologous recombination. Preferably, one or more inactivating mutations, such as a stop codon or frame-shift mutation, are also introduced in the OCH1 sequence in the vector to prevent the production of any potentially active OCH1 polypeptide.

In still another aspect, the present invention provides methods of producing a glycoprotein having a mammalian-like N-glycan structure. A nucleotide sequence coding for a glycoprotein of interest can be introduced into a methylotrophic yeast strain which has been engineered to produce mammalian-like N-glycans. Alternatively, a methylotrophic yeast strain which expresses a glycoprotein of interest can be modified to express the desired enzymes (i.e., α-1,2-mannosidase, GnTI and GalT) and to inactivate the genomic OCH1 gene, in order to produce the glycoprotein with mammalian-like N-glycans.

In still another aspect, glycoproteins produced by using the methods of the present invention, i.e., glycoproteins having mammalian-like N-glycans, particularly the $GalGlcNAcMan_5GlcNAc_2$ N-glycan, are provided by the present invention.

In a further aspect, the present invention provides a kit containing one or more of the vectors of the present invention, or one or more of the genetically engineered strains of the present invention.

Brief Description of the Drawings

Figure 1 depicts the structures of $M_8GlcNAc_2$, $M_5GlcNAc_2$, $GlcNAcM_5GlcNAc_2$, and Gal $GlcNAcM_5GlcNAc_2$.

Figure 2 graphically depicts yeast and human N-linked glycosylation and the strategy for humanization of the *Pichia pastoris* glycosylation. The glyco-engineering steps include inactivation of the α-1,6-mannosyltransferase *OCH1*, overexpression of a HDEL tagged α-1,2-mannosidase and Golgi-localized GnTI and GalT. The final partially obtained hybrid structure is framed.

Figure 3A graphically depicts the strategy for inactivating the genomic OCH1 gene by single homologous recombination.

Figure 3B graphically depicts plasmid pGlycoSwitchM5 used for glycan engineering of *Pichia pastoris*. Upon linearization of pGlycoSwitchM5 with *Bst* BI, subsequent transformation and correct integration in the genome of *P. pastoris*, the *OCH1* gene was inactivated.

Figure 3C graphically depicts pPIC6AKrecoGnTI.

Figure 3D graphically depicts pBlKanMX4KrehGalT.

Figure 4 graphically depicts DSA-FACE analysis of N-glycans from different glycan engineered *Pichia pastoris* strains. Panel 1: Oligomaltose reference. Panels 2-9 represent N-glycans from – 2: wild type strain GS115, with $Man_9GlcNAc_2$ representing the main peak; 3: *och1* inactivated strain, with $Man_8GlcNAc_2$ representing the main peak; 4: *och1* inactivated ManHDEL expressing strain, with $Man_5GlcNAc_2$ representing the main peak; 5: *och1* inactivated ManHDEL, KreGnTI expressing strain, with $GlcNAcMan_5GlcNAc_2$ representing the main peak; 6: same as 5 except that glycans were treated with β-N-acetylhexosaminidase, and the $GlcNAcMan_5GlcNAc_2$ peak shifted to the $Man_5GlcNAc_2$ position, indicating that terminal GlcNAc was present; 7: *och1* inactivated ManHDEL, KreGnTI, KreGalT expressing strain, with the additional peak representing $GalGlcNAcMan_5GlcNAc_2$, which disappeared when treated with β-galactosidase; 9: reference glycans from bovine RNase B ($Man_{5-9}GlcNAc_2$).

Figures 5A-5B demonstrate glycosylation after inactivation of *Pichia pastoris OCH1*. 5A: CBB stained SDS-PAGE gel of supernatant of *T. reesei* mannosidase secreting *Pichia pastoris* strains. In the non-engineered strain (WT) a clear smear was visible whereas this smear was absent in the *och1* inactivated strain (och1 (M8)). 5B: FACE analysis of N-glycans derived from mannosidase secreted by a non-engineered strain (WT) and an *och1* strain. The bands with higher electrophoretic mobility are indicated with Man8 and Man9 and represent "core" N-glycan structures.

Detailed Description of the Invention

The present invention is directed to methods, vectors and genetically engineered methylotrophic yeast strains for making recombinant glycoproteins with mammalian-like or human-like glycosylation.

By "mammalian" is meant to include any species of mammal, such as human, mice, cats, dogs, rabbits, cattle, sheep, horse and the like.

Typical complex type mammalian glycans, such as glycans produced in humans, have two to six outer branches with a sialyl-N-acetyl-lactosamine sequence linked to an inner core structure of $Man_3GlcNAc_2$. Mammalian N-glycans originate from a core oligosaccharide structure, $Man_8GlcNAc_2$, which is formed in the ER. Proteins with this core sugar structure are transported to the Golgi apparatus where $Man_8GlcNAc_2$ is converted to $Man_5GlcNAc_2$ as a result of the removal of 3 mannose residues by Golgi mannosidases I (Golgi α-1,2-mannosidases). As proteins proceed through the Golgi, $Man_5GlcNAc_2$ is further modified by the addition of GlcNAc and the removal of two more mannose residues, followed by the addition of GlcNAc, galactose (Gal), and sialic acid (SA) residues.

The term "mammalian-like glycosylation" as used herein is meant that the N-glycans of glycoproteins produced in a genetically engineered methylotrophic yeast strain include five or fewer mannose residues and are characteristic of N-glycans or intermediate carbohydrate structures in the biosynthesis of N-glycans of proteins, produced in mammalian cells such as human cells.

In a preferred embodiment, glycoproteins produced in a genetically engineered methylotrophic yeast strain of the present invention include five or fewer mannose residues, and at least one N-acetylglucosamine residue (GlcNAc) linked to the core structure containing mannose residues, and to a terminal galactose residue. For example, glycoproteins produced in a genetically engineered methylotrophic yeast strain have $GalGlcNAcMan_5GlcNAc_2$, as graphically depicted in **Figure 1**. The IUPAC nomenclature of this carbohydrate ($GalGlcNAcMan_5GlcNAc_2$) is Gal(β-1,4)GlcNAc(β-1,2)Man(α-1,3){ Man(α-1,3) [Man(α-1,6)] Man(α-1,6)}Man(β-1,4)GlcNAc(β -1,4)GlcNAc. Its extended nomenclature is β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→2)-α-D-Man*p*-(1→3)-{α-D-Man*p*-(1→3)-[α-D-Man*p*-(1→6)]-α-D-Man*p*-(1→6)}-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-D-Glc*p*NAc.

It has been established that the majority of N-glycans on glycoproteins leaving the endoplasmic reticulum (ER) of methylotrophic yeasts, including *Pichia* and especially *Pichia pastoris*, have the $Man_8GlcNAc_2$ oligosaccharide structure. After the glycoproteins are transported from the ER to the Golgi apparatus, additional mannose residues are added to this core sugar moiety by different mannosyltransferases, resulting in glycoproteins with oligosaccharide structures consisting of a high manose core, or extended, branched mannan outer chains.

According to the present invention, in order to produce recombinant glycoproteins with mammalian-like glycosylation, methylotrophic yeasts are modified to express the enzymes that convert the carbohydrate structure, $Man_8GlcNAc_2$, in a series of steps to mammalian-like N-glycans. Preferably, methylotrophic yeasts are also modified to inactivate the expression of one or more enzymes involved in the production of high mannose structures, e.g., α-1,6-mannosyltransferase encoded by the OCH1 gene.

The term "methylotrophic yeast" as used herein includes, but is not limited to, yeast strains capable of growing on methanol, such as yeasts of the genera *Candida, Hansenula, Torulopsis,* and *Pichia.* Preferred methylotrophic yeasts of the present invention are strains of the genus *Pichia.* Especially preferred are *Pichia pastoris* strains

GS115 (NRRL Y-15851), GS190 (NRRL Y-18014), PPF1 (NRRL Y-18017), PPY120H, YGC4, and strains derived therefrom.

In one embodiment, the present invention provides a genetically engineered methylotrophic yeast strain which produces glycoproteins having a mammalian-like N-glycan structure, characterized as having five or fewer mannose residues and at least one N-acetylglucosamine residue (GlcNAc) which is linked to the core mannose-containing structure and to a terminal galactose residue.

In a preferred embodiment, the present invention provides a genetically engineered methylotrophic yeast strain which produces glycoproteins having the mammalian-like N-glycan structure, $GalGlcNAcMan_5GlcNAc_2$.

According to the present invention, the methylotrophic yeast strain which produces glycoproteins having $GalGlcNAcMan_5GlcNAc_2$ is genetically engineered to express an α-1,2-mannosidase or a functional part thereof, an N-acetylglucosaminyltransferase I (or GnTI) or a functional part thereof, and a β-1,4-galactosyltransferase (GalT) or a functional part thereof. Preferably, the methylotrophic yeast strain is also genetically engineered such that the genomic OCH1 gene is inactivated.

An α-1,2-mannosidase cleaves the α-1,2-linked mannose residues at the non-reducing ends of $Man_8GlcNAc_2$, and converts this core oligosaccharide on glycoproteins to $Man_5GlcNAc_2$, which is the acceptor substrate for the mammalian N-acetylglucosaminyltransferase I.

According to the present invention, a methylotrophic yeast strain can be engineered to express an α-1,2-mannosidase or a functional part thereof by introducing into the strain, e.g., by transformation, a nucleotide sequence encoding the α-1,2-mannosidase or the functional part thereof. The nucleotide sequence encoding an α-1,2-mannosidase or a functional part thereof can derive from any species. A number of α-1,2-mannosidase genes have been cloned and are available to those skilled in the art, including mammalian genes encoding, e.g., a murine α-1,2-mannosidase (Herscovics et al. *J. Biol. Chem.* 269: 9864-9871, 1994), a rabbit α-1,2-mannosidase (Lal et al. *J. Biol.*

Chem. 269: 9-872-9881, 1994) or a human α-1,2-mannosidase (Tremblay et al. *Glycobiology* 8: 585-595, 1998), as well as fungal genes encoding, e.g., an *Aspergillus* α-1,2-mannosidase (msdS gene), or a *Trichoderma reesei* α-1,2-mannosidase (Maras et al. *J. Biotechnol.* 77: 255-263, 2000. Protein sequence analysis has revealed a high degree of conservation among the eukaryotic α-1,2-mannosidases identified so far.

Preferably, the nucleotide sequence for use in the present vectors encodes a fungal α-1,2-mannosidase, more preferably, a *Trichoderma reesei* α-1,2-mannosidase, and more particularly, the *Trichoderma reesei* α-1,2-mannosidase described by Maras et al. *J. Biotechnol.* 77: 255-63 (2000).

By "functional part" is meant a polypeptide fragment of an α-1,2-mannosidase which substantially retains the enzymatic activity of the full-length protein. By "substantially" is meant at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length α-1,2-mannosidase is retained. Characterizations of various domains, including the catalytic domain, of a number of α-1,2-mannosidases are documented. See, e.g., "Isolation of a mouse Golgi mannosidase cDNA, a member of a gene family conserved from yeast to mammals", Herscovics et al., *J Biol Chem* 269:13 9864-71 (1994); "Isolation and expression of murine and rabbit cDNAs encoding an alpha 1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides", Lal et al., *J Biol Chem* 269:13 9872-81 (1994); "Molecular cloning and enzymatic characterization of a Trichoderma reesei 1,2-alpha-D-mannosidase", Maras M et al., *J Biotechnol* 77:255-63 (2000); and U.S. Patent Application 20020188109, incorporated herein by reference. Those skilled in the art can also readily identify and make functional parts of an α-1,2-mannosidase using a combination of techniques known in the art. The activity of a portion of an α-1,2-mannosidase of interest, expressed and purified from an appropriate expression system, can be verified using *in vitro* or *in vivo* assays described in U.S. Patent Application 20020188109, incorporated herein by reference.

In accordance with the present invention, an α-1,2-mannosidase or a functional part thereof expressed in a methylotrophic yeast strain preferably is targeted to a site in the secretory pathway where $Man_8GlcNAc_2$ (the substrate of α-1,2-mannosidase)

is already formed on a glycoprotein, but has not reached a Golgi glycosyltransferase which elongates the sugar chain with additional mannose residues. In a preferred embodiment of the present invention, the α-1,2-mannosidase or a functional part thereof is engineered to contain an ER-retention signal such that the α-1,2-mannosidase or a functional part thereof, which is expressed in the methylotrophic yeast strain is targeted to the ER.

"An ER retention signal" refers to a peptide sequence which directs a protein having such peptide sequence to be transported to and retained in the ER. Such ER retention sequences are often found in proteins that reside and function in the ER. Multiple choices of ER retention signals are available to those skilled in the art, e.g., the first 21 amino acid residues of the *S. cerevisiae* ER protein MNS1 (Martinet et al. *Biotechnology Letters* 20: 1171-1177, 1998), and the peptide HDEL (SEQ ID NO: 1).

A preferred ER retention signal for use in the present invention is the peptide HDEL (SEQ ID NO: 1). The HDEL peptide sequence, which is found in the C-terminus of a number of yeast proteins, acts as a retention/retrieval signal for the ER (Pelham *EMBO J.* 7: 913-918, 1988). Proteins with an HDEL sequence are bound by a membrane-bound receptor (Erd2p) and then enter a retrograde transport pathway for return to the ER from the Golgi apparatus.

The α-1,2-mannosidase for use in the present invention can be further engineered, e.g., to contain an epitope tag to which antibodies are available, such as Myc, HA, FLAG and His6 tags well-known in the art. An epitope-tagged α-1,2-mannosidase can be conveniently purified, or monitored for both expression and intracellular localization.

According to the present invention, an ER retention signal can be placed, by genetic engineering, anywhere in the protein sequence of an α-1,2-mannosidase, but preferably at the C-terminus of the α-1,2-mannosidase.

An ER retention signal and an epitope tag can be readily introduced into an α-1,2-mannosidase or a functional part thereof by inserting a nucleotide sequence coding

for such signal or tag into the nucleotide sequence encoding the α-1,2-mannosidase or the functional part, using any of the molecular biology techniques known in the art.

The expression of an α-1,2-mannosidase in an engineered methylotrophic yeast strain can be verified both at the mRNA level, e.g., by Northern Blot analysis, and at the protein level, e.g., by Western Blot analysis. The intracellular localization of the protein can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence experiments. The localization of an α-1,2-mannosidase in the ER can be determined by co-sedimentation of this enzyme with a known ER resident protein (e.g., Protein Disulfide Isomerase) in a subcellular fractionation experiment. The localization in the ER can also be determined by an immunofluorescence staining pattern characteristic of ER resident proteins, typically a perinuclear staining pattern.

To confirm that an α-1,2-mannosidase or a functional part thereof expressed in a methylotrophic yeast strain has the expected mannose-trimming activity, both *in vitro* and *in vivo* assays can be employed. Typically, an *in vitro* assay involves digestion of an *in vitro* synthesized substrate, e.g., $Man_8GlcNAc_2$, with the enzyme expressed and purified from a methylotrophic yeast strain, and assessing the ability of such enzyme to trim $Man_8GlcNAc_2$ to, e.g., $Man_5GlcNAc_2$. In *in vivo* assays, the α-1,2-mannosidase or a part thereof is co-expressed in a methylotrophic yeast with a glycoprotein known to be glycosylated with N-glycans bearing terminal α-1,2-linked mannose residues in such yeast. The enzymatic activity of such an α-1,2-mannosidase or a part thereof can be measured based on the reduction of the number of α-1,2-linked mannose residues in the structures of the N-glycans of the glycoprotein. In both *in vitro* and *in vivo* assays, the composition of a carbohydrate group can be determined using techniques that are well known in the art and are illustrated in the Examples hereinbelow.

Further according to the present invention, a methylotrophic yeast strain can be engineered to express a GlcNAc-Transferase I or a functional part thereof by introducing into the strain, e.g., by transformation, a nucleotide sequence encoding the GlcNAc-Transferase I or the functional part thereof. A GlcNAc-Transferase I is

responsible for the addition of β-1,2-GlcNAc to a $Man_5GlcNAc_2$, and converts this core oligosaccharide on glycoproteins to $GlcNAcMan_5GlcNAc_2$. The mannose residues of $GlcNAcMan_5GlcNAc_2$ can be further trimmed by a mammalian Golgi mannosidase II, and additional sugar units, such as galactose, can be added towards forming hybrid- or complex-type sugar branches characteristic of mammalian glycoproteins.

The nucleotide sequence encoding a GlcNAc-transferase I (GnTI) or a functional part thereof for introduction into a methylotrophic yeast strain can derive from any species, e.g., rabbit, rat, human, plants, insects, nematodes and protozoa such as *Leishmania tarentolae*. Preferably, the nucleotide sequence for use in the present invention encodes a human GnTI, and more preferably, the human GnTI as set forth in SEQ ID NO: 13.

By "functional part" of a GnTI is meant a polypeptide fragment of the GnTI, which substantially retains the enzymatic activity of the full-length GnTI. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length GnTI is retained. The enzymatic activity of a GnTI or a portion thereof can be determined by assays described in Reeves et al. (*Proc. Natl. Acad. Sci. U S A.* 99(21):13419-24, 2002), Maras et al. (*Eur J Biochem.* 249 (3):701-7, 1997), or in the Examples hereinbelow. Those skilled in the art can readily identify and make functional parts of a GnTI using a combination of techniques known in the art. For example, as illustrated by the present invention, the catalytic domain (containing the last 327 residues) of the human GnTI constitutes a "functional part" of the human GnTI.

In accordance with the present invention, a GnTI or a functional part thereof expressed in a methylotrophic yeast strain is preferably targeted to a site in the secretory pathway where $Man_5GlcNAc_2$ (the substrate of GnTI) is already formed on a glycoprotein. Preferably, the GnTI or a functional part thereof is targeted to the Golgi apparatus.

Accordingly, in a preferred embodiment of the present invention, the GnTI or a functional part thereof is engineered to contain a Golgi localization signal.

A "Golgi localization signal" as used herein refers to a peptide sequence, which directs a protein having such sequence to the Golgi apparatus of a methylotrophic

yeast strain and retains the protein therein. Such Golgi localization sequences are often found in proteins that reside and function in the Golgi apparatus.

Choices of Golgi localization signals are available to those skilled in the art. A preferred Golgi localization signal for use in the present invention is a peptide derived from the N-terminal part of a *Saccharomyces cerevisiae* Kre2 protein (ScKre2); more preferably, the ScKre2 protein as set forth in SEQ ID NO: 10. A particularly preferred Golgi localization signal is the peptide (SEQ ID NO: 11), composed of amino acids 1-100 of the ScKre2 protein as set forth in SEQ ID NO: 10.

According to the present invention, a Golgi localization signal can be placed anywhere within a GnTI, but preferably at the terminus of the GnTI, and more preferably at the N-terminus of the GnTI.

The GnTI for use in the present invention can be further engineered, e.g., to contain an epitope tag to which antibodies are available, such as Myc, HA, FLAG and His6 tags, which are well-known in the art. An epitope-tagged GnTI can be conveniently purified, or monitored for both expression and intracellular localization.

A Golgi localization signal and an epitope tag can be readily introduced into a GnTI by inserting a nucleotide sequence coding for such signal or tag into the nucleotide sequence encoding the GnTI, using any of the molecular biology techniques known in the art.

Further according to the present invention, a methylotrophic yeast strain can be engineered to express a β-1,4-galactosyltransferase (GalT) of a functional part thereof by introducing into the strain, typically by transformation, a nucleotide sequence encoding the a β-1,4-galactosyltransferase (GalT) of the functional part thereof. GalT adds a β-1-4-galactose residue to the GlcNAc on the left arm of the glycan structure ($GlcNAcMan_5GlcNAc_2$), as depicted in **Figure 1**.

The nucleotide sequence encoding a GalT or a functional part thereof for introduction into a methylotrophic yeast strain can derive from any species, e.g. mammalians (e.g. humans, mice), plants (e.g. *Arabidopsis thaliana*), insects (e.g. *Drosophila melanogaster*), or nematodes (e.g. *Caenorhabditis elegans*). Preferably, the

nucleotide sequence for use in the present invention encodes a human GalT, and more preferably, the human GalT1 as set forth in SEQ ID NO: 21.

By "functional part" of a GalT is meant a polypeptide fragment of the GalT, which substantially retains the enzymatic activity of the full-length GalT. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length GalT is retained. The enzymatic activity of a GalT or a portion thereof can be determined by assays described in Maras et al. (*Eur J Biochem.* 249(3):701-7, 1997) or in the Examples hereinbelow. Those skilled in the art can readily identify and make functional parts of a GalT using a combination of techniques known in the art. For example, as illustrated by the present invention, the catalytic domain of the human GalT constitutes a "functional part" of the human GalT.

In accordance with the present invention, a GalT or a functional part thereof expressed in a methylotrophic yeast strain is preferably targeted to a site in the secretory pathway where $GlcNAcMan_5GlcNAc_2$ (a substrate of GalT) is already formed on a glycoprotein. Preferably, the GalT or a functional part thereof is targeted to the Golgi apparatus.

Accordingly, in a preferred embodiment of the present invention, the GalT or a functional part thereof is engineered to contain a Golgi localization signal as described hereinabove. A preferred Golgi localization signal for targeting a GalT to the Golgi apparatus is the peptide (SEQ ID NO: 11), composed of amino acids 1-100 of the ScKre2 protein as set forth in SEQ ID NO: 10.

According to the present invention, a Golgi localization signal can be placed anywhere within a GalT, but preferably at the terminus of the GalT, and more preferably at the N-terminus of the GalT.

The GalT for use in the present invention can be further engineered, e.g., to contain an epitope tag to which antibodies are available, such as Myc, HA, FLAG and His6 tags, well-known in the art. An epitope-tagged GalT can be conveniently purified, or monitored for both expression and intracellular localization.

A Golgi localization signal and an epitope tag can be readily introduced into a GalT by inserting a nucleotide sequence coding for such signal or tag into the nucleotide

sequence encoding the GalT, using any of the molecular biology techniques known in the art.

To achieve expression of a desirable protein (i.e., an α-1,2-mannosidase, a GnTI, a GalT, or a functional part of any of these enzymes) in a methylotrophic yeast strain, the nucleotide sequence coding for the protein can be placed in a vector in an operable linkage to a promoter and a 3' termination sequence that are functional in the methylotrophic yeast strain. The vector is then introduced into the methylotrophic yeast strain, e.g., by transformation.

Promoters appropriate for expression of a protein in methylotrophic yeast include both constitutive promoters and inducible promoters. Constitutive promoters include e.g., the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase promoter ("the GAP promoter"). Examples of inducible promoters include, e.g., the *Pichia pastoris* alcohol oxidase I promoter ("the AOXI promoter") (U.S. Patent No. 4,855,231), or the *Pichia pastoris* formaldehyde dehydrogenase promoter ("the FLD promoter") (Shen et al. *Gene* 216: 93-102, 1998).

3' termination sequences are sequences 3' to the stop codon of a structural gene which function to stabilize the mRNA transcription product of the gene to which the sequence is operably linked, such as sequences which elicit polyadenylation. 3' termination sequences can be obtained from *Pichia* or other methylotrophic yeasts. Examples of *Pichia pastoris* 3' termination sequences useful for the practice of the present invention include termination sequences from the *AOX1* gene and the *HIS4* gene.

Transformation of vectors or linear fragments thereof can be achieved using any of the known methods, such as the spheroplast technique, described by Cregg et al. (*Mol. Cell. Biol.* (12): 3376-85, 1985), or the whole-cell lithium chloride yeast transformation system, described by Ito et al. (*Agric. Biol. Chem. 48(2)*:341, (1984)), modified for use in *Pichia* as described in EP 312,934. Other methods useful for transformation include those described in U.S. Patent No. 4,929,555; Hinnen et al. (*Proc. Nat. Acad. Sci. USA* 75:1929 (1978)); Ito et al. (*J. Bacteriol.* 153:163 (1983)); U.S. Patent No. 4,879,231; and Sreekrishna et al. (*Gene* 59:115 (1987)). Electroporation and PEG1000 whole cell transformation procedures can also be used. See Cregg and Russel,

Methods in Molecular Biology: Pichia Protocols, Chapter 3, Humana Press, Totowa, N.J., pp. 27-39 (1998).

Transformed yeast cells can be selected by using appropriate techniques including but not limited to culturing auxotrophic cells after transformation in the absence of the biochemical product required (due to the cell's auxotrophy), selection for and detection of a new phenotype, or culturing in the presence of an antibiotic which is toxic to the yeast in the absence of a resistance gene contained in the transformants. Transformants can also be selected and/or verified by integration of the expression cassette into the genome, which can be assessed by e.g., Southern Blot or PCR analysis.

As described hereinabove, in addition to expression of an α-1,2-mannosidase, and N-acetylglucosaminyltransferase I (or GnTI), a β-1,4-galactosyltransferase (GalT), or a functional part thereof, the methylotrophic yeast strain is preferably also genetically engineered to inactivate the genomic OCH1 gene in order to efficiently produce glycoproteins having the $GalGlcNAcMan_5GlcNAc_2$ glycan.

The OCH1 gene encodes a membrane bound α-1,6-mannosyltransferase that is localized in the early Golgi complex and initiates the α-1,6-polymannose outer chain addition to the N-linked core oligosaccharide ($Man_5GlcNAc_2$ and $Man_8GlcNAc_2$). The *S. cerevisiae* OCH1 gene and a *Pichia* OCH1 gene have been cloned (Nakayama et al. *EMBO J.* 11: 2511-2519, 1992, and Japanese Patent Application No. 07145005, respectively). Those skilled in the art can isolate the OCH1 genes from other methylotrophic yeasts using techniques well known in the art.

According to the present invention, a disruption of the OCH1 gene of a methylotrophic yeast strain can result in either the production of an inactive protein product or no product. The disruption may take the form of an insertion of a heterologous DNA sequence into the coding sequence and/or the deletion of some or all of the coding sequence. Gene disruptions can be generated by homologous recombination essentially as described by Rothstein (in *Methods in Enzymology*, Wu et al., eds., vol 101:202-211, 1983).

To disrupt the genomic OCH1 gene by double homologous recombination, an OCH1 "knock-out" vector can be constructed, which includes a selectable marker gene, operably linked at both its 5' and 3' ends to portions of the OCH1 gene of lengths sufficient to mediate homologous recombination. The selectable marker can be one of any number of genes which either complement host cell auxotrophy or provide antibiotic resistance, including *URA3*, *ARG4, HIS4, ADE1*, *LEU2 HIS3, Sh ble* (*Streptoalloteichus hindustanus* bleomycin gene) and *BSD* (blasticidin S deaminase from *Aspergillus terreus*) genes. Other suitable selectable markers include the invertase gene from *Saccharomyces cerevisiae*, which allows methylotrophic yeasts to grow on sucrose; or the *lacZ* gene, which results in blue colonies due to the expression of active β-galactosidase. A linear DNA fragment of an OCH1 inactivation vector, which contains the selectable marker gene with OCH1 sequences at both its 5' and 3' end, is then introduced into host methylotrophic yeast cells using any of the transformation methods well known in the art. Integration of the linear fragment into the genomic OCH1 locus and the disruption of the OCH1 gene can be determined based on the selection marker and can be verified by, for example, Southern Blot analysis.

Alternatively, an OCH1 knock-out vector can be constructed which includes a portion of the OCH1 gene, wherein the portion is devoid of any OCH1 promoter sequence and encodes none or an inactive fragment of the OCH1 protein. By "an inactive fragment" is meant a fragment of the full-length OCH1 protein, which fragment has, preferably, less than about 10%, and more preferably, about 0% of the activity of the full-length OCH1 protein. Such portion of the OCH1 gene is inserted in a vector with no operably linkage to any promoter sequence that is functional in methylotrophic yeast. This vector can be subsequently linearized at a site within the OCH1 sequence, and transformed into a methylotrophic yeast strain using any of the transformation methods known in the art. By way of single homologous recombination, this linearized vector is then integrated in the OCH1 locus, resulting in two och1 sequences in the chromosome, neither of which is able to produce an active Och1p protein, as depicted in **Figure 3A**.

Preferably, an inactivating mutation is also introduced in the och1 sequence in the vector at a site 5' to (upstream of) the linearization site and 3' to (downstream of) the

translation initiation codon of OCH1. By "inactivating mutation" is meant a mutation that introduces a stop codon, a frameshift mutation or any other mutation causing a disruption of the reading frame. Such mutation can be introduced into an och1 sequence in a vector using any of the site directed mutagenesis methods known in the art. Such inactivating mutation ensures that no functional Och1p protein is formed after homologous recombination, even if there exist some promoter sequences 5' to the Och1 sequence in the knock-out vector.

The genetically engineered methylotrophic yeast strains, as described hereinabove, can be further modified if desired. For example, disruption of additional genes encoding any other *Pichia* mannosyltransferases can be made. Genes encoding enzymes that function in the mammalian glycosylation pathway, other than α-1,2-mannosidase, GnTI or GalT, can be introduced to increase the proportion of mammalian-like N-glycans and/or to further modify the mammalian-like N-glycans, if desired. For example, the genetically engineered methylotrophic yeast strains described above can be further modified to express the *S. cerevisiae GAL10*-encoded enzyme, which converts UDP-glucose into UDP-galactose and vice versa. This may increase the level of cytosolic UDP-galactose, which then stimulates the activity of GalT and increase the proportion of the $GalGlcNAcM_5GlcNAc_2$ glycans. In addition, the genetically engineered methylotrophic yeast strains described above can be further modified to express a mannosidase II in the Golgi, which removes additional mannose residues from $GalGlcNAcM_5GlcNAc_2$ thereby permitting addition of other sugar residues.

The sequence of the genetic modifications is not critical to the present invention. Introduction of nucleotide sequences encoding an α-1,2-mannosidase, a GnTI and a GalT, and disruption of the genomic OCH1 gene, can be conducted sequentially, in any order, or simultaneously by co-transformation with two or more different vectors or coding sequences or by transformation with one vector which include two or more different coding sequences.

In a further aspect, the present invention provides vectors useful for generating methylotrophic yeast strains which produce glycoproteins having a

mammalian-like N-glycan structure, characterized as having five or fewer mannose residues and at least one N-acetylglucosamine residue (GlcNAc) which is linked to the core mannose-containing structure and to a terminal galactose residue, e.g., $GalGlcNAcMan_5GlcNAc_2$.

In one embodiment, the present invention provides a vector which contains a nucleotide sequence coding for an enzyme to be expressed, i.e., an α-1,2-mannosidase, a GnTI, a GalT, or a functional part of any of these proteins. Such vectors are also referred to as "knock-in" vectors. The coding sequence can be placed in an operable linkage to a promoter and a 3' termination sequence that are functional in the host methylotrophic yeast for expression of the encoded protein. Two or more coding sequences can be placed in the same vector for simultaneous transformation into a methylotrophic yeast strain. Preferably, the vector also includes any one of the selectable marker gene as described hereinabove for convenient selection of transformants.

According to the present invention, the knock-in vectors, which contain a sequence coding for a desirable protein to be expressed in a methylotrophic yeast strain, can be either an integrative vector or a replicative vector (such as a replicating circular plasmid). Integrative vectors are disclosed, e.g., in U.S. Patent No. 4,882,279, which is incorporated herein by reference. Integrative vectors generally include a serially arranged sequence of at least a first insertable DNA fragment, a selectable marker gene, and a second insertable DNA fragment. The first and second insertable DNA fragments each can be about 200 nucleotides in length and have nucleotide sequences which are homologous to portions of the genomic DNA of the species to be transformed. A nucleotide sequence containing a structural gene of interest for expression is inserted in this vector between the first and second insertable DNA fragments whether before or after the marker gene. Integrative vectors can be linearized prior to yeast transformation to facilitate the integration of the nucleotide sequence of interest into the host cell genome.

In another embodiment, the present invention provides an inactivation vector (or a "knock-out" vector) which, when introduced into a methylotrophic yeast strain, inactivates or disrupts the genomic OCH1 gene.

The vector for inactivating genomic OCH1 gene can include a selectable marker gene, which is operably linked, at both its 5' and 3' end, to portions of the OCH1 gene of lengths sufficient to mediate homologous recombination, as described hereinabove. Transformation of methylotrophic yeast cells with a linear DNA fragment of such an OCH1 inactivation vector, which contains the selectable marker gene with OCH1 sequences at both its 5' and 3' end, leads to integration of the linear fragment into the genomic OCH1 locus and disruption of the genomic OCH1 gene.

Alternatively, an OCH1 inactivation vector can include a portion of the OCH1 gene to be disrupted, which portion encodes none or an inactive fragment of the OCH1 protein, and any one of the selectable marker gene as described hereinabove. Such portion of the OCH1 gene is devoid of any OCH1 promoter sequence and is not in an operable linkage to any known promoter sequence. Such vector can be linearized at a site within the Och1 sequence and subsequently transformed into a methylotrophic yeast strain, which results in inactivation of the genomic OCH1 gene by a single homologous recombination-mediated integration. Preferably, an inactivating mutation, such as a stop codon or frame-shift mutation, is also introduced in the Och1 sequence in the vector at a site 5' to (upstream of) the linearization site and 3' to (downstream of) the translation initiation codon of OCH1.

If desired, a nucleotide sequence coding for an enzyme to be expressed in a methylotrophic yeast strain can be combined with a nucleotide sequence capable of inactivating the genomic OCH1 gene, in the same vector to create a "knock-in-and-knock-out" vector.

The vectors of the present invention, including both knock-in vectors and knock-out vectors, can also contain selectable marker genes which function in bacteria, as well as sequences responsible for replication and extrachromosomal maintenance in bacteria. Examples of bacterial selectable marker genes include ampicillin resistance (*Amp*^{r}), tetracycline resistance (*Tet*^{r}), hygromycin resistance, blasticidin resistence and zeocin resistance (*Zeo*^{R}) genes.

Additionally, any of the above-described vectors can further include a nucleotide sequence encoding a glycoprotein of interest for expression of such glycoprotein in a methylotrophic yeast strain.

In still another aspect, the present invention provides methods of producing a glycoprotein having a mammalian-like N-glycan structure.

"A glycoprotein" as used herein refers to a protein which, in methylotrophic yeasts, is either glycosylated on one or more asparagines residues or on one or more serine or threonine residues, or on both asparagines and serine or threonine residues. Preferably, the glycoprotein is heterologous to the host methylotrophic yeast strain.

In accordance with the present invention, the production of a glycoprotein of interest with reduced glycosylation can be achieved in a number of ways. For example, a nucleotide sequence coding for a glycoprotein of interest can be introduced into a methylotrophic yeast strain which has been previously engineered to produce mammalian-like N-glycans.

The nucleotide sequence coding for a glycoprotein can be placed in an operably linkage to a promoter sequence and a 3' termination sequence that are functional in the host strain. The nucleotide sequence can include additional sequences, e.g., signal sequences coding for transit peptides when secretion of a protein product is desired. Such signal sequences are widely known, readily available and include *Saccharomyces cerevisiae* alpha mating factor prepro(αmf), the *Pichia pastoris* acid phosphatase (PHO1) signal sequence and the like.

Alternatively, a methylotrophic yeast strain which has been introduced with a coding sequence for a glycoprotein of interest, can be modified to express the desired enzymes (i.e., a-1,2-mannosidase, GnTI and GalT) and to inactivate the genomic OCH1 gene, as described hereinabove, in order to produce the glycoprotein having mammalian-like N-glycans.

Glycoproteins produced in methylotrophic yeasts can be purified by conventional methods. Purification protocols can be determined by the nature of the specific protein to be purified. Such determination is within the ordinary level of skill in the art. For example, the cell culture medium is separated from the cells and the protein

secreted from the cells can be isolated from the medium by routine isolation techniques such as precipitation, immunoadsorption, fractionation or a variety of chromatographic methods.

Glycoproteins which can be produced by the methods of the present invention include bacterial, fungal or viral proteins or antigens, e.g., *Bacillus amyloliquefaciens* α-amylase, *S. cerevisiae* invertase, *Trypanosoma cruzi trans*-sialidase, HIV envelope protein, influenza virus A haemagglutinin, influenza neuraminidase, Bovine herpes virus type-1 glycoprotein D; proteins, a protein of a mammalian origin, such as human proteins, growth factors or receptors, e.g., human angiostatin, human B7-1, B7-2 and B-7 receptor CTLA-4, human tissue factor, growth factors (e.g., platelet-derived growth factor), tissue plasminogen activator, plasminogen activator inhibitor-I, urokinase, human lysosomal proteins such as α-galactosidase, plasminogen, thrombin, factor XIII; and immunoglobulins or fragments (e.g., Fab, Fab', $F(ab')_2$) of immunoglobulins. For additional useful glycoproteins which can be expressed in the genetically engineered *Pichia* strains of the present invention, see Bretthauer and Castellino, *Biotechnol. Appl. Biochem.* 30: 193-200 (1999), and Kukuruzinska et al., *Ann Rev. Biochem.* 56: 915-944 (1987).

Glycoproteins produced by using the methods of the present invention, i.e., glycoproteins having mammalian-like N-glycans, particularly the $GalGlcNAcMan_5GlcNAc_2$ N-glycan, are also part of the present invention.

In still another aspect, the present invention provides a kit which contains one or more of the knock-in vectors, knock-out vectors, or knock-in-and-knock-out vectors of the present invention described above.

More particularly, a kit of the present invention contains a vector having a nucleotide sequence coding for an α-mannosidase I or a functional part thereof, preferably containing an ER-rentention signal; a vector having a nucleotide sequence coding for a GnTI or a functional part thereof, preferably containing a Golgi-rentention signal; a vector having a nucleotide sequence coding for a GalT or a functional part

thereof, preferably containing a Golgi-rentention signal; or a vector capable of disrupting the genomic OCH1 gene in a methylotrophic yeast, or any combinations thereof.

The kit can also include a nucleic acid molecule having a sequence coding for a heterologous glycoprotein of interest. Such nucleic acid molecule can be provided in a separate vector or in the same vector which contains sequences for knocking-in or knocking out as described hereinabove. Alternatively, the knock-in or knock-out vectors in the kit have convenient cloning sites for insertion of a nucleotide sequence encoding a heterologous protein of interest.

The kit can also include a methylotrophic yeast strain which can be transformed with any of the knock-in, knock-out or knock-in-and-knock-out vectors described hereinabove. Alternatively, the kit can include a methylotrophic yeast strain which has been engineered to produce mammalian-like N-glycans.

The present invention is further illustrated by the following examples.

Example 1

Materials And Methods

Vector Construction And Transformation

A *Pichia pastoris* sequence was found in the GenBank under Accession No. E12456 (SEQ ID NO: 2) and was described in Japanese Patent Application No. 07145005, incorporated herein by reference. This sequence shows all typical features of an α-1,6-mannosyltransferase and is most homologous to the *S. cerevisiae* OCH1, thus referred to herein as the *Pichia pastoris* OCH1 gene.

The full ORF of the *Pichia pastoris* OCH1 gene was isolated by PCR using genomic DNA isolated from strain GS115 as template and the following oligonucleotides: 5'GGAATTCAGCATGGAGTATGGATCATGGAGTCCGTTGGAAAGG (SEQ ID NO: 4), and 5'GCCGCTCGAGCTAGCTTTCTTTAGTCC (SEQ ID NO: 5). The isolated OCH1 gene was cloned in pUC18 to obtain plasmid pUC18pOCH1, and the identity of the OCH1 gene sequence was confirmed by sequencing.

Plasmid pGlycoSwitchM8 (2875 bp, SEQ ID NO: 6, graphically depicted in **Figure 3A**) contains a fragment of the *Pichia pastoris* OCH1 ORF encoding Ala25-Ala155, which fragment was inserted between the *Bgl* II and *Hind* III sites of pPICZB (Invitrogen, Carlsbad, CA). Two stop-codons were situated in frame just before codon Ala25 to prevent the possible synthesis of a truncated protein. The *Bst*B I site of the polylinker of pPICZB was previously eliminated by filling in and religation after digestion. The unique *Bst*B I site located inside the cloned *OCH1* fragment can be used for linearization of the plasmid (See **Figure 3A** for an overview of the inactivation strategy).

pGlycoSwitch M5 (5485 bp, SEQ ID NO: 9, graphically depicted in **Figure 3B**) was constructed as follows. An *Xba* I / *Cla* I fragment of pPIC9 (Invitrogen, Carlsbad, CA), containing the *Pichia pastoris* HIS4 transcriptional terminator sequence, was inserted between the *Hind* III and *Eco*R I sites of pGlycoSwitch M8. Afterwards the 2.3 kb *Bgl* II / *Not* I fragment of pGAPZMFManHDEL (Callewaert et al., *FEBS Lett*, 503(2-3):173-178, 2001) containing the GAP promoter and preMFmannosidaseHDEL

cassette, was inserted between the *Hind* III and *Not* I sites. All restriction sites used for this construction (except for the *Not* I site) were filled in with Klenow DNA polymerase. The unique *Bst*B I site in pGAPZMFmanHDEL was previously eliminated by filling and religation after digestion.

In order to target the human GlcNAc-transferase I (GnTI) to the Golgi apparatus, the GnTI N-terminal part was replaced by the *S. cerevisiae* Kre2 N-terminal part that is responsible for the localization in the yeast Golgi (Lussier et al., *J Cell Biol*, 131(4):913-927, 1995). Plasmid YEp352Kre2 (provided by Dr. Howard Bussey, McGill University, Montreal, Canada) was generated by inserting the *Sac* I/*Pvu* II fragment of the Kre2 gene in the Yep352 vector, which vector had been digested with *Sal* I (blunted with Klenow) and *Sac* I. YEp352Kre2 was digested with *Sac* I/*Pvu* I and made blunt by T4-polymerase. The 5'end of the Kre2 gene was isolated and cloned in a Klenow blunted *Sgr*A I / *Xba* I opened pUChGnTI (Maras et al., *Eur J Biochem* 249(3):701-707, 1997). The fusion place between the two DNA fragments was sequenced using standard procedures. The resulting Kre2-GnTI open reading frame that contained the N-terminal part of the Kre2 gene (encoding the first 100 amino acids of the Kre2 protein, as set forth in SEQ ID NO: 11) and the catalytic domain of GnTI (the last 327 amino acids of GnTI which is as set forth in SEQ ID NO:13) was isolated by an *Eco*R V / *Hind* III double digest and ligated in a *Sal* I / *Eco*R I opened pPIC6A vector (Invitrogen) after blunting of both fragments with Klenow polymerase. The resulting plasmid was named pPIC6AKrecoGnTI (SEQ ID NO: 14, graphically depicted in **Figure 3C**). It contains the Kre2GnTI open reading frame under control of the methanol inducible AOX1 promotor and *BSD* gene from *A. terreus* for resistance against the antibiotic blasticidin.

Localization of GalT was achieved by fusion of the catalytic domain of GalT to the N-terminal part of Kre2p in the same way as was done to target GnTI. β-1,4-galactosyltransferase was amplified from a hepg2 cDNA library using oligonucletides 5'TTCGAAGCTTCGCTAGCTCGGTGTCCCGATGTC (SEQ ID NO: 15) and 5'GAATTCGAAGGGAAGATGAGGCTTCGGGAGCC (SEQ ID NO: 16) as starter sequences. The amplified fragment was cloned *Hind* III / *Eco*R I into pUC18. To omit the N-terminal 77 amino acids of the GalT protein, a PCR was performed using the

following oligonucleotides as primers:
5'TTCGAAGCTTCGCTAGCTCGGTGTCCCGATGTC (SEQ ID NO: 15) and
5'CGTTCGCGACCGGAGGGGCCCGGCCGCC (SEQ ID NO: 17). The amplified fragment was cut with *Nru* I / *Hind* III and ligated into the *Hind* III / *Sgr*A I Klenow blunted pUCKreGnTI vector. The resulting Kre2-GalT fusion construct was again amplified by PCR using the as primers:
5'TCGATATCAAGCTTAGCTCGGTGTCCCGATGTC (SEQ ID NO: 18) and
5'GAATTCGAACTTAAGATGGCCCTCTTTCTCAGTAAG (SEQ ID NO: 19). The amplified fragment was cloned *Eco*R V / *Bst*B I into the pBLURA IX (Cereghino et al., *Gene*, 263:159-169, 2001) (provided by James Cregg, Oregon Graduate Institute of Science and Technology, Beaverton, USA). Finally the *URA3* gene was replaced by a Kanamycin resistance cassette by ligating a *Spe* I / *Sma* I fragment from the vector pFA6a-KanMX4 into the *Spe* I / *Ssp* I opened plasmid. The final plasmid, named as pBlKanMX4KrehGalT (SEQ ID NO: 22, graphically depicted in **Figure 3D**), contained the sequence encoding a Kre2-GalT fusion protein, operably linked to the AOX1 promoter. The fusion protein was composed of the first 100 amino acids of Kre2 and the last 320 amino acids of GalT.

Transformations of these plasmids to GS115 *Pichia* strains expressing various proteins were performed as described previously (Cregg et al., *Methods in Molecular Biology*, 103:27-39, 1998). Correct genomic integration at the Pp*OCH1* locus was confirmed by PCR on genomic DNA.

Protein preparation

Secreted *Trichoderma reesei* α-1,2-mannosidase was purified using a combination of HIC, anion exchange and gel filtration chromatography, as described (Maras et al., J Biotechnol, 77(2-3):255-263, 2000; Van Petegem et al., *J Mol Biol* 312(1):157-165, 2001). All SDS-PAGE experiments were done on 10% PAA gels under standard running conditions. Yeast cell wall mannoproteins were released as described by Ballou (*Methods Enzymol*, 185:440-470, 1990), which involved extensive washing of

yeast cells with 0.9% NaCl in water, prolonged autoclavation of the yeast cells (90 min) in 20mM Na-citrate after, followed by methanol precipitation (4 volumes).

N-glycan analysis

N glycan analysis was conducted by laser-induced DNA-sequencer assisted fluorophore-assisted carbohydrate electrophoresis on the ABI 377 DNA-sequencer (DSA-FACE), as described (Callewaert et al., *Glycobiology*, 11(4):275-281, 2001). In short, glycoproteins were immobilized on a Multiscreen Immobilon-P plate and deglycosylated by PNGase treatment. N-glycans were recovered and derivatized with APTS. Excess of label was removed by size fractionation on a Sephadex G10 resin. After evaporation of the APTS-labeled oligosaccharides, a ROX-labeled GENESCAN 500 standard mixture (Applied Biosystems) was added to allow internal standardization. This mixture was run on an ABI 377A DNA sequencer (Applied Biosystems) with a 12% polyacrylamide gel in an 89 mM Tris, 89 mM borate, 2.2 mM EDTA buffer. On each gel, N-glycans of bovine RNase B and a maltodextrose ladder was run as a reference. Data analysis was performed using the GENESCAN 3.1 software (Applied Biosystems). Exoglycosidase treatment with β-N-acetylhexosaminidase (Glyko) and β-galactosidase (Prozyme), was performed on labeled glycans overnight at 37°C in 20 mM sodium acetate pH 5.5. Conventional FACE (ANTS labeling of N-glycans and electrophoresis on 30% PAA mini gels) was performed as described by Jackson (*Biochem J*, 270(3):705-713, 1990). The DSA-FACE method had a very high resolution and sensitivity, while the conventional FACE was well suited for detecting complex mixtures of higher molecular weight N-glycans ('hyperglycosylation'), which were not resolved and therefore formed a characteristic 'smear' on the gel in conventional FACE. Thus, a combination of DSA-FACE and conventional FACE analyses gave a more complete picture of the characteristics of yeast-produced glycoproteins.

Growth curve determination

The fresh overnight yeast cultures were diluted with fresh YPD medium to OD600 0.02 and grown overnight at 250 rpm, 30°C (12 hours, OD 600 < 3.0). To start

the experiment, 10 mL of fresh YPD in 50 mL polypropylene tubes were inoculated with overnight yeast cultures to get starting an OD600 value of 0.5. Aliquotes were taken every 2 hours and OD600 values were measured. All yeast strains were run at the same time in parallel.

Example 2

Inactivation of OCH1

Disruption of the genomic *Pichia pastoris* OCH1 gene was achieved by single homologous recombination as follows. The plasmid, pGlycoSwitchM8 (**Figure 3A**), was generated as described in Example 1, which included base pairs No. 73-467 of the *Pichia pastoris* OCH1gene, preceded by two in-frame non-sense codons to avoid read-through from potential earlier translation start sites in the vector. This fragment contained a centrally located *Bst*B I site useful for linearization of the vector before transformation, and was linked at its 3' end to the AOX1 transcription terminator sequence. This vector would duplicate the OCH1 sequence present in the vector upon integration by single homologous recombination into the genomic OCH1 locus of *Pichia*. As a result, the OCH1 gene in the *Pichia* chromosome was replaced with two Och1 sequences. The first OCH1 sequence encoded a protein product of 161 amino acids long at maximum (of which 6 amino acids resulted the from the sequence in the vector), which did not include the catalytic domain of the type II transmembrane protein encoded by the full-length OCH1 gene. The second OCH1 sequence lacked the coding sequence for the first 25 amino acids of the full-length protein, and contained two in-frame stop codons that would prevent any read-through from potential upstream translation initiation sites.

Strain GS115 was transformed with the plasmid pGlycoSwitchM8. The transformant was referred to as GlycoSwitchM8 or, in short, the M8 strain or the *och1* strain. PCR on genomic DNA with the primer combinations specified in **Figure 3A**, showed correct integration of this construct in the expected genomic locus in about 50% of Zeocin resistant transformants, as indicated by three independent experiments.

Analysis of the cell wall mannoprotein N-glycans revealed a change in glycosylation pattern as can be deduced from **Figure 4**. Whereas the predominant peak

is $Man_9GlcNAc_2$ for the cell wall mannoprotein from the wild type GS115 strain, the main peak is $Man_8GlcNAc_2$ for the GlycoSwitchM8 strain (compare panels 2 to 3 of **Figure 4**). This change in N-glycans was reverted after transformation of the M8 strain with the full-length *OCH1* ORF.

To evaluate whether the heterogeneity of secreted glycoproteins from the M8 strain was decreased, *T. reesei* α-1,2-mannosidase, which is a typically hyperglycosylated, secreted protein in the wild type GS115 strain (Maras et al., *J Biotechnol*, 77(2-3):255-263, 2000), was analyzed using the *och1* M8 strain. The culture supernatant of cells of the M8 strain, which had been transformed with a nucleotide sequence coding for *T. reesei* α-1,2-mannosidase, was separated by SDS-PAGE (**Figure 5A**). The gel reveals that the smear, characteristic of hyperglycosylated proteins, was absent in the proteins produced in the GlycoSwitchM8 strain. In parallel, the secreted glycoproteins were deglycosylated by the PNGase F treatment, and the glycans were analyzed by FACE analysis on mini-gels. Typically in FACE analysis, large hyperglycosyl structures are not resolved and appear as one smearing band (**Figure 5B**). The smearing band was absent with glycoproteins from the *och1* strain, confirming that the heterogeneity of the N-glycans from the *och1* strain was decreased.

Example 3

Expression of ER retained mannosidase-HDEL

To further humanize the N-glycans of *Pichia pastoris*, ER retained *Trichoderma reesei* α-1,2-mannosidase-HDEL was expressed in the *och1* strain. For easy conversion of a *Pichia pastoris* expression strain, a nucleotide sequence coding for *Trichoderma reesei* α-1,2-mannosidase-HDEL was inserted into the *och1* inactivation vector. The resulting combination vector was called pGlycoSwitchM5, the construction of which is described in Example 1.

Strain GS115 was transformed with linearized pGlycoSwitchM5. Correct integration of the vector was confirmed by PCR analysis. N-glycans of mannoproteins

from the transformants were analyzed by the DSA-FACE method. The glycan profile revealed a homogenous $Man_5GlcNAc_2$ peak (**Figure 4**, panel 4). Integration of the $Man_5GlcNAc_2$ peak and of all the small peaks above the detection limit of this method (S/N>3) in the size area of 5 up to 25 glucose units revealed that this higher-eukaryote type high-mannose glycan made up for at least 90% of the total N-glycan pool present in this mixture.

In an alternative approach, the mannosidase-HDEL was expressed under control of the methanol inducible AOX1 promoter. No apparent differences in N-glycan profile between the two mannosidase-expressing strains (*i.e.* constitutive and inducible) could be detected.

To confirm the N-glycan modifications of a heterologous protein, the pGlycoSwitchM5 plasmid was transformed into a *Trypanosoma cruzi trans*-sialidase expressing *Pichia* strain as described by Laroy et al. (*Protein Expr Purif*, 20(3):389-393, 2000). Here too, $Man_5GlcNAc_2$ was detected on the purified protein, accounting for more than 95% of total N-glycan on the purified protein.

Growth curve analysis of the pGlycoSwitchM5 transformed strain in shake flask culture indicated that its doubling time closely mimicked that of the wild type strain. However, the engineered strain reached the stationary phase at an optical density that was about 20% lower than the wild type strain, indicating that it could be somewhat more sensitive to the stress conditions of high cell density. Nevertheless, its stress sensitivity phenotype was much less pronounced than the *S. cerevisiae och1* strain.

Example 4

Expression of Golgi-localized N-acetylglucosaminyltransferase I (Kre2GnTI)

To target GnTI to the Golgi, the nucleotide sequence coding for the N-terminal part of GnTI, including the cytosolic part, the transmembrane region and a part of the luminal stem region, was replaced with a nucleotide sequence coding for the *S. cerevisiae* Kre2 signal sequence. This resulted in a nucleotide sequence coding for a

chimeric protein having the first 100 amino acids from Kre2p and the last 327 amino acids of GnTI.

For expression in *Pichia pastoris*, the Kre2-GnTI chimeric sequence was placed under control of the strong methanol inducible AOX1 promoter in a plasmid having the blasticidin resistance marker. The resulting construct, pPIC6KrecoGnTI (as described in Example 1), was transformed into a GS115 M5 strain after linearization in the *AOX1* locus by digestion with *Nsi* I. The presence of the construct in the transformants was confirmed by PCR on genomic DNA using AOX1 3' and 5' primers.

N-glycans of mannoproteins of several transformants were analyzed by the DSA-FACE method. The dominant peak was about one glucose unit larger than the $Man_5GlcNAc_2$ peak (**Figure 4**, panel 5). To determine whether this peak had terminal GlcNAc, an exoglycosidase digest was performed with β-N-acetylhexosaminidase, an enzyme that hydrolyzes β-GlcNAc linkages. Upon digestion with this enzyme, the peak shifted back to the $Man_5GlcNAc_2$ (**Figure 4**, panel 6). This indicates that the original peak represents $GlcNAcMan_5GlcNAc_2$, and thus confirms the correct *in vivo* activity of the chimeric GnTI enzyme.

Overexpression of the Kre2GnTI chimer led to an almost complete conversion of $Man_5GlcNAc_2$ to $GlcNAcMan_5GlcNAc_2$. This suggests that enough UDP-GlcNAc donor substrate was present in the Golgi to N-acetylglucosaminylate almost all the N-glycans.

Example 5

Expression of Golgi retained β-1,4-galactosyltransferase

The nucleotide sequence coding for the N-terminal part of human β-1,4-galactosyltransferase 1 (the first 77 amino acids), including the transmembrane domain and the cytosolic part of the enzyme, was replaced by a nucleotide sequence coding for the *S. cerevisiae* Kre2 signal sequence. This chimeric fusion sequence was placed under control of the AOX1 promotor and the 3' end of AOX1 as a terminator. The final

plasmid, pBlKanMX4KrehGalT (described in Example 1), was linearized with *Pme* I prior to transformation into the M5-GnTI strain.

N-glycan analysis was done with mannoproteins from several transformants. A peak about one glucose unit larger than the $GlcNAcMan_5GlcNAc_2$ peak was detected in the transformants, whereas the peak was absent in the non-transformed strain (**Figure 3**, panel 7). The N-glycans were digested with β-galactosidase to determine whether this peak represented glycans containing terminal β-galactose. After digestion of the glycan profile, this peak shifted back to the $GlcNAcMan_5GlcNAc_2$ position (**Figure 4**, panel 8 in comparison to panel 7). The amount of $GalGlcNAcMan_5GlcNAc_2$ was determined by integrating the $GlcNAcMan_5GlcNAc_2$ peak before and after the β-galactosidase digestion. Subtraction of these two peaks revealed that about 10% of $GlcNAcMan_5GlcNAc_2$ was converted to $GalGlcNAcMan_5GlcNAc_2$. Supplementing the medium with 0.2% galactose did not increase the amount of Gal-containing oligosaccharides.